Surprise Minimizing Multi-Agent Learning with Energy-Based Models

Karush Suri **Xiao Qi Shi** **Kostas Plataniotis** **Yuri Lawryshyn**

@karush_

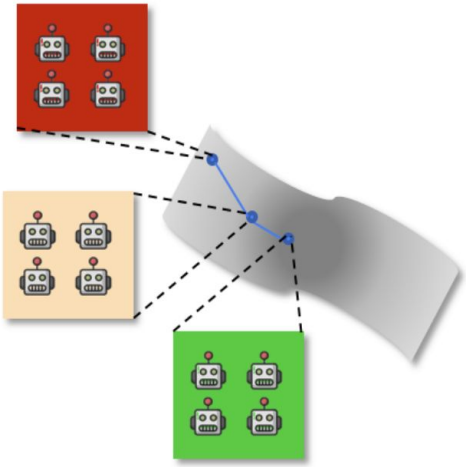
karush17.github.io/emix-web

Surprise in Multi-Agent Learning

- **Sudden Environmental changes impact agent behavior**
- **Uncertainty due to abrupt temporal changes**
- **Surprising states grow with the number of agents**
- **What if we treat surprise as energy and minimize it?**

An Energy-Based Approach

- **Map surprise to an energy landscape and seek the minima**



- **Utilise energy operator as a contraction on the surprise value function**

Theorem 1. *Given a surprise value function* $V^a_{\text{surp}}(s,u,\sigma)\ \forall a \in N$, *the energy operator* $\mathcal{T}V^a_{\text{surp}}(s,u,\sigma) = \log \sum_{a=1}^{N} \exp\left(V^a_{\text{surp}}(s,u,\sigma)\right)$ *forms a contraction on* $V^a_{\text{surp}}(s,u,\sigma)$.

- **Minimize surprise via intrinsic motivation**

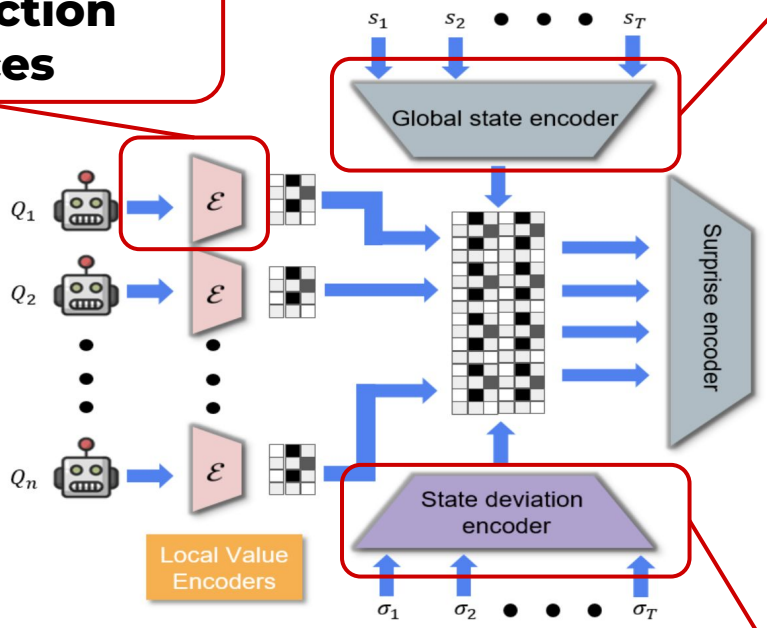
$$\hat{Q}(u,s;\theta) = Q(u,s;\theta) + \beta \log \sum_{a=1}^{N} \exp\left(V^a_{\text{surp}}(s,u,\sigma)\right))$$

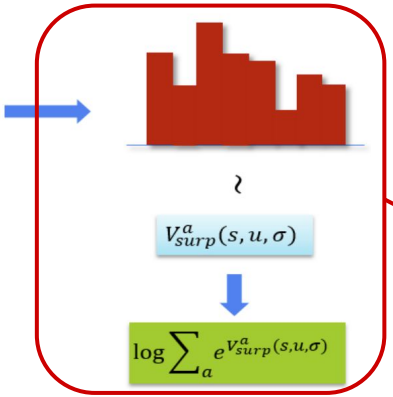
Penalize Q values if high energy (surprise)

Energy-Based Surprise Minimization: EMIX

Local encoders for large action spaces

Latent state representation

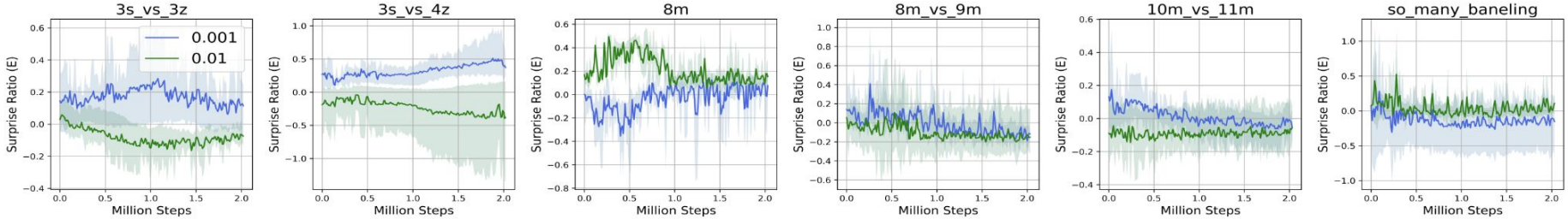




Estimate & minimize energy

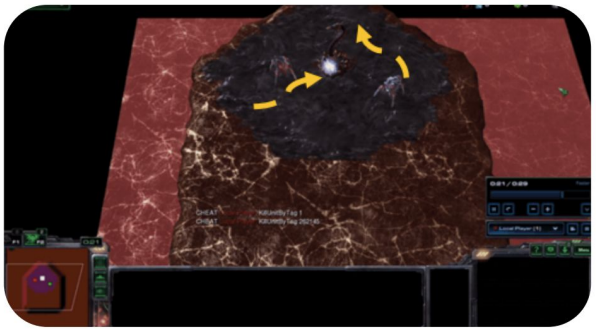
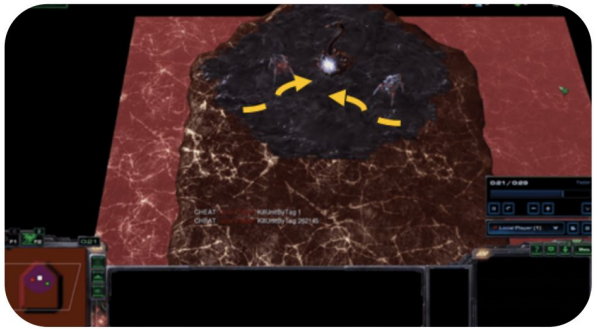
Encoding uncertainty

Key Results



Scenarios	EMIX	SMiRL-QMIX	QMIX	VDN	COMA	IQL
3m	**94.90±0.39**	93.94±0.22	93.43±0.20	94.58±0.58	84.75±7.93	94.79±0.50
3s_vs_4z	**97.22±0.73**	0.24±0.11	96.01±3.93	94.29±2.13	0.00±0.00	59.75±12.22
8m_vs_9m	**71.03±2.69**	69.90±1.94	**68.28±2.30**	58.81±4.68	4.17±0.58	28.48±22.38
10m_vs_11m	75.35±2.30	**77.85±2.02**	70.36±2.87	71.81±6.50	4.55±0.73	32.27±25.68
so_many_baneling	**95.87±0.16**	93.61±0.94	93.35±0.78	92.26±1.06	91.65±2.26	74.97±6.52
5m_vs_6m	**37.07±2.42**	33.27±2.79	34.42±2.63	35.63±3.32	0.52±0.13	14.78±2.72

Qualitative Analysis

EMIX	QMIX
	
	

Thank You

@karush_

karush17.github.io/emix-web